Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 2 (2010) 129–136

Procedia
Social and Behavioral Sciences

Security Camera Network, Privacy Protection and Community Safety

Background image generation by preserving lighting condition of outdoor scenes

Yasutomo Kawanishi[a,]*, Ikuhisa Mitsugami[b], Masayuki Mukunoki[b], Michihiko Minoh[b]

[a]*Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Kyoto 606-8501, Japan*

Received November 13, 2009; revised December 4, 2009; accepted December 10, 2009

Abstract

In the Sensing Web project, we have proposed the concept of “Henshin” camera, which outputs only privacy-protected information in order to open and share the information on the Internet. The information should not contain appearances of people in the captured scene but just their positions and the background image of the scene. In this paper, a novel method is proposed to generate the verisimilar background image which well expresses the weather and the lighting condition of the scene. This method collects a huge number of images by super long term surveillance, classifies them according to their time, and applies the eigenspace method so as to reproduce the background image without any appearances of the people in it. We experimentally evaluate the results of this approach using data of several surveillance cameras.

Keywords: Background generation; super long-term surveillance; eigenspace method; “Henshin” camera; privacy protection

1. Introduction

In these days, many surveillance cameras are installed in our daily space for several purposes; traffic surveillance, security, weather forecast, etc. Each of these cameras is used only for its purpose and the captured video is used only by its installer. The video is, however, thought to have also other information than that for the original purpose. If the video is shared among many persons through the Internet, the camera is expected to be utilized more conveniently and effectively. For example, we could get weather information from a traffic surveillance camera, congestion degree of shoppers in a shopping mall from a security camera, and so on. Considering these usages, we notice the usefulness of opening and sharing real-time sensory information on the Internet. The Sensing Web project (Minoh, 2009) (Minoh, Kakusho, Babaguchi, & Ajisaka, 2008), which was launched under this consideration, proposes to connect all available sensors including the cameras to the Internet, and to open the sensory data for many persons so that anyone can use the real-time sensory data for various purposes from anywhere.

On opening and sharing the sensory data, the most serious problem is privacy invasion of observed people. As long as the sensory data is closed in a certain system operated by an institution like most existing systems, the

* Corresponding author. Tel.: +81-75-753-9065; fax: +81-75-753-9056.
E-mail address: ykawani@mm.media.kyoto-u.ac.jp.

1877-0428
10.1016/j.sbspro.2010.01.026

privacy information can be managed and controlled by the corresponding institution so that we do not need to take care of the problem. On the other hand, in the case of the sensing web, the sensory data is opened to the public so that anyone can access any sensory data without any access managements. Especially, the video, which is the sensory data obtained by the cameras, contains rich information of the people and may cause the privacy invasion. In fact, a person in a video can be easily identified by his/her appearance features (face, motion, colors of cloths, etc.). The privacy invasion problem, therefore, has been a main obstacle against opening the sensory data.

In the sensing web project, we tackle the problem to realize an infrastructure where any sensory data is opened and shared. To overcome the problem, the privacy information has to be erased from the image. One of the ways to realize this privacy elimination is to mask the appearances of the people on the image. In fact, Google Street View (GSV) (Google, 2008) adopts this approach. This service faces the same problem as mentioned above, although it offers not real-time sensory data but just the snapshots at a past moment. In GSV, each person in the captured image is detected and masked automatically. This operation can be executed using a human detection technique. However, as the technique does not works perfectly, some people cannot masked correctly so that their privacy cannot be protected; when a person is detected in a wrong position or not detected, the mask is overlaid on a wrong position or is not overlaid so that the person is left unmasked and clearly appeared in the output image. We thus propose another approach to overcome the privacy invasion problem based on a noble idea; the image of the camera is reconstructed by generating a background image without any people and overlaying symbols at the positions of the corresponding people on the generated image. This idea is implemented as a “Henshin” camera (Fig.1.). In the case of this “Henshin” camera, even if the human detection does not work well, it just causes the rendering of the symbol on the different position or the lack of the character, but never causes the privacy invasion.

Fig.1. The outputs of a traditional security camera and “Henshin” camera

For realizing this “Henshin” camera, we need two techniques; a human detection and a background image generation. The former one has been studied for a long time, and there are a lot of existing studies. They are mainly categorized into two types of approaches; background subtraction methods (Haritaoglu, Harwood, & Davis, 1999) (Jabri, Duric, Wechsler, & Rosenfeld, 2000) and human detection methods (Chen, & Chen, 2008) (Dalal, & Triggs, 2005). In this paper, we use a HoG-based human detection (Dalal, & Triggs, 2005), which is known as a method that works robustly even when the luminosity of the scene changes frequently. On the other hand, the latter one has to be well considered. Although it looks just a conventional issue at a glance, it is indeed much different from the many existing methods for the background generations. Considering the concept of the “Henshin” camera, we have to design the background generation method ensuring that people would never appear in the output image even if a person stops for quite a long time in the scene, which is treated as not the foreground person but the background by the most methods. Besides, our method has to generate the background image as verisimilar to the truth as possible, because we would like to know the real-time weather of the observed area by this background image. Especially, lighting condition by the sun is helpful to know the weather so that such kind of information has to be well reconstructed. Considering these discussions above, this paper proposes a novel background generation technique which preserves the shadow accurately in outdoor scene while ensuring that a person never appears in the image. This technique is realized by collecting the images for super long term, categorizing them by time, and analyzing them using the eigenspace method. The detail of the technique is mentioned in the following sections.

2. Traditional Background Generation Methods

If a human detection performs perfectly so that all the people in the image can be erased, the whole image except the people regions can be used as the accurate background image. However, when the people exist, the corresponding regions would be left as blanks so that the background image cannot be always fully generated. In addition, there has been no ideal method for the human detection, which is apparent to see that there are still many challenges for this topic. Therefore, in terms of the privacy protection, we must not directly use each image which is observed by camera. We have to take an analytic approach collecting many images for a certain term.

Calculating median or average of each pixel of the image sequence is a simple approach to generate the background image generation. In the case of this kind of approach, in order to follow the background changing, short term image sequence is used. However, people who stop for the term appear in the generated image, which causes the privacy invasion. For generating the background avoiding privacy invasion, we have to analyze images collected by much longer term than people might stop. On the other hand, in terms of reconstructing the background image as similar to the truth as possible especially from the viewpoint of the lighting condition by the sun, such analytic approaches using long term image sequence do not perform well; they cannot follow immediately the sudden and frequent changes of the strength of the sunlight, because the generated image is influenced by the many images in the past. Such approaches cannot fulfill the demand for applying to the "Henshin" camera.

The eigenspace method is often used to analyze huge amount of data. Recently, the method is applied to analyze the tendency of huge amount of images. We apply this method to the images collected by long term surveillance. Using the eigenspace method (Oliver, Rosario, & Pentland, 1999), we can analytically reconstruct the background image from the current image captured by the camera that may contain some people. This is achieved by the following process.

First they calculate the eigenvectors $e_1, e_2, \cdots, e_k$ (sorted in descending order by their contributions) from a number of images by the principal component analysis (PCA). As the eigenspace defined by these eigenvectors indicates the variation of the image sequence, the background image x_t^b can be estimated from observed image x_t using this eigenspace; x_t^b is calculated by the following equation:

$$x_t^b \approx Ep = EE^T x_t,$$

where p describes the corresponding point in the eigenspace and $E = [e_1 \quad e_2 \quad \cdots \quad e_k]$ is an orthonormal matrix.

In this method, each image is expected not to contain any people, because we want to extract the tendency of only the background without influenced by the appearances of the people. As mentioned repeatedly, we have no way to judge perfectly whether there are any people or not. We thus have to use the images each of which may contain people. Note that the appearance of the people should have less influence than the variation of the background, as the people are usually much smaller than the size of the image and each of them moves randomly and is observed for just a short term. Thus, even if we use such images, we can get the eigenspace which includes no influence of people by using only s $(s << k)$ eigenvectors to reconstruct the background. We use the orthonormal matrix $E' = [e_1 \quad e_2 \quad \cdots \quad e_s]$ instead of E to estimate the background image x_t^b (Fig.2). x_t^b is calculated by the following equation:

$$x_t^b \approx E'p' = E'E'^T x_t$$

where p' describes the corresponding point in the eigenspace. It means that even when we use the image which may contain people, we expect to get the similar result to the case using the images without any people in them when we choose only such the small number of the eigenvectors.

Nevertheless, another problem still exist; the lighting condition may not be able to reconstructed by such the small (s) dimensional eigenspace.

3. Eigenspace Method with Classification by Observed Time

In outdoor scene, there are sharp shadow edges in the observed images and the position of the shadow edges are shifted gradually caused by solar position. When we generate the background which includes various shadows in the scene, we need eigenvectors in the very high dimensional eigenspace which correspond to each position of shadow edges. As discussed in the previous section, when the dimension s is small, such the eigenvectors corresponding to the moving shadow edges may be neglected, so that we could not generate the background image keeping such various shadow edges by linear sum of top s eigenvectors (Fig.3).

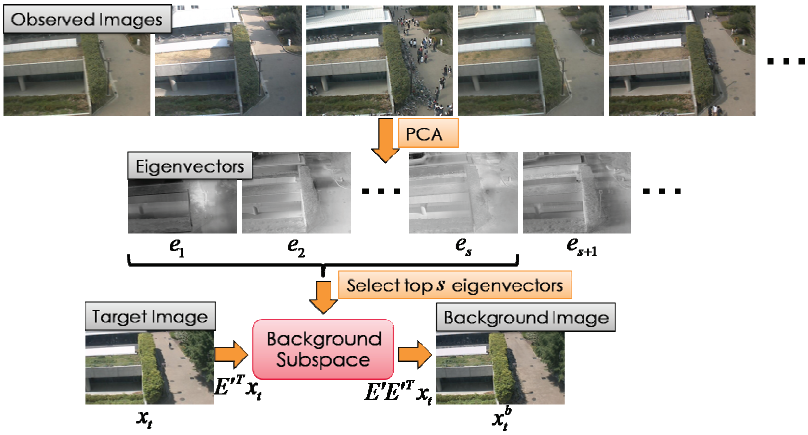


Fig.2. The background generation by the eigenspace method using observed images including foreground objects.

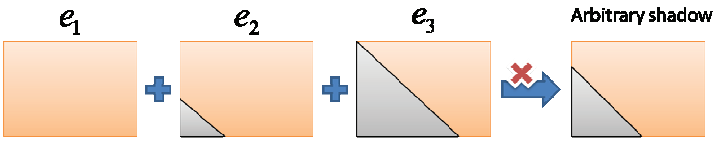


Fig.3. Various shadow edges cannot describe as linear sum of the small number of the eigenvectors.

Our method relies on the fact that the shadows appear in the similar position in the same time even in another day. We collect huge amount of images with super long term surveillance and classify them into image sets according to their capture time. The images of each set are expected to have similar spatial appearance of shadows. Fig.4. shows the shadow appearances. Looking at the images observed in a day, there are shadows in the different position. On the other hand, we can see that the shadows appear in the similar position when we look at the images observed in 15 o'clock of different days.

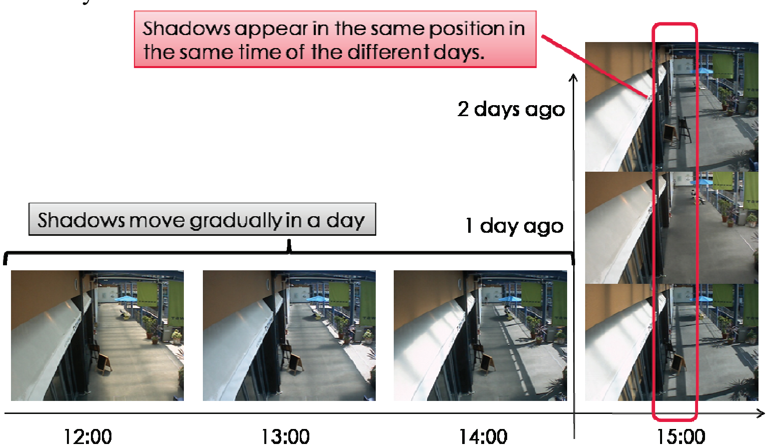
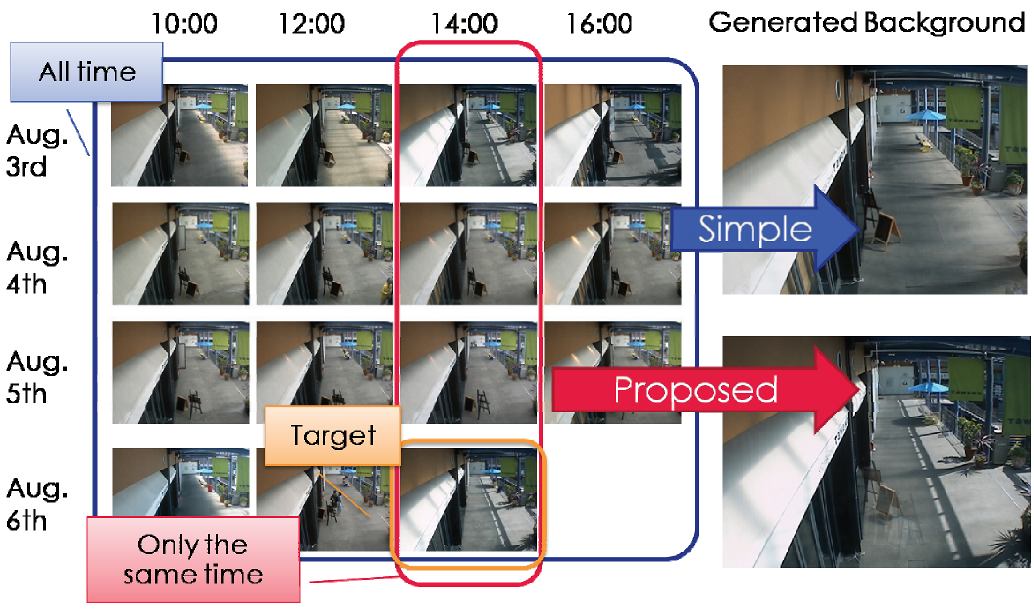
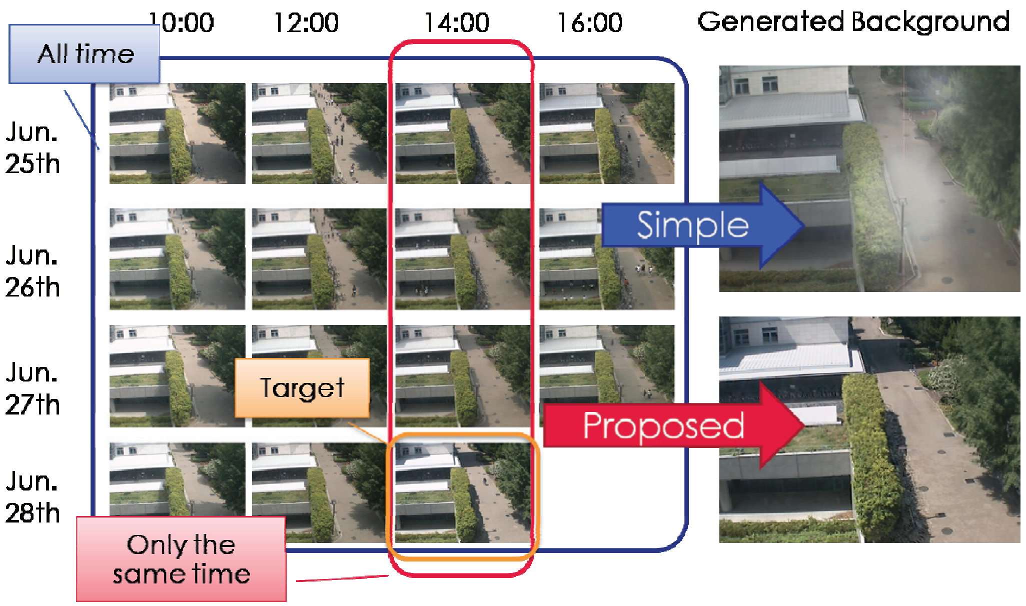


Fig.4. Examples of the shadow positions. Shadow moves gradually in a day, and appear in the similar position in the images which were observed in the same time of the different days.

Thus, the procedure of the proposed method is as follow. First, at a time t of a day d, we get a target image $x_{d,t}$. Then, we classify the observed images $x_{d,t}$ by the observed time t, and we can get an image set $x_{d-1,t}, x_{d-2,t}, \cdots$ which were observed in the time t of the different days $d-1, d-2, \cdots$. We describe the image set as I_t. Finally, we apply the eigenspace method to the image set I_t, and then we can get the background keeping lighting conditions for each target image $x_{d,t}$ which is a raw image that may contain some people in it.



(a) Case 1 : an example of the background generations



(b) Case 2: another example of the background generations

Fig.5. Examples of the background image generation. We applied both the simple eigenspace method and the proposed method to the image observed at (a) 14:00 Aug.3rd and (b) 14:00 Jun.28th, and estimate the background images. We generated the 4-dimensional eigenspaces for images observed in 1 month. For the proposed method, we used images observed in 15 seconds around 14:00 of everyday. The shadows appeared clearly in the result of our method, while the shadows do not appear clearly in the result of the previous method.

4. Evaluation

To show the effectiveness of our method, we experimented in some outdoor scenes. We generated the background images by the simple eigenspace method and the proposed method. We used the following two scenes:

- Scene 1: The input images and the results are shown in Fig.5 (a). These images are observed in August 1^{st} – 31^{st} at a shopping mall. We try to generate the background image of a target image observed at 14:00 in August 6^{th}.
- Scene 2: The input images and the results are shown in Fig5 (b). These images are observed in June 1^{st} – 30^{th} at a shopping mall. We try to generate the background image of a target image observed at 14:00 in August 28^{th}.

For each scene, we applied two methods: the simple eigenspace method which uses all the images for calculating the eigenspace, and the proposed method which first classifies the images into the sets according to the time and then applies the eigenspace method to the set. For the proposed method, we used images observed in 15 seconds around 14:00 of everyday. We used 4-dimensional eigenspace, which is considered to appropriate for these methods by another experiment mentioned in the following paragraph. Comparing those two results in each scene, it is visually confirmed that the proposed method performed better from the viewpoint of keeping the lighting condition.

To determine the most appropriate number of dimensions, we performed another experiment. For each dimensional eigenspace from 1-dimensional to 30-dimensional one, we generated the background images of the input images, and differential images between the input images and the generated background images. Then the averages of the pixel values for each differential image, which is graphically shown in Fig.6, are calculated. The horizontal axis of the graph indicates the dimension of the eigenspaces and the vertical one indicates the average of the pixel values of the differential images. When the number of dimension is smaller than around five, the average decreases because the eigenspace does not have enough information of the illuminative variation of the background. On the other hand, when the number of dimension becomes larger than around five, the average gets increased. This is because ghosts of objects which is not moved in a day but re-located at another position in another day like signage boards become to be seen in the generated images. Considering the graph, we selected the most appropriate number of the dimension as the minimum point of the graph of each method; 5-dimensional eigenspace for the simple eigenspace method and 4-dimensional eigenspace for the proposed one.

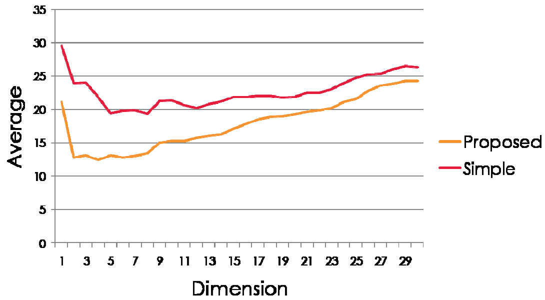
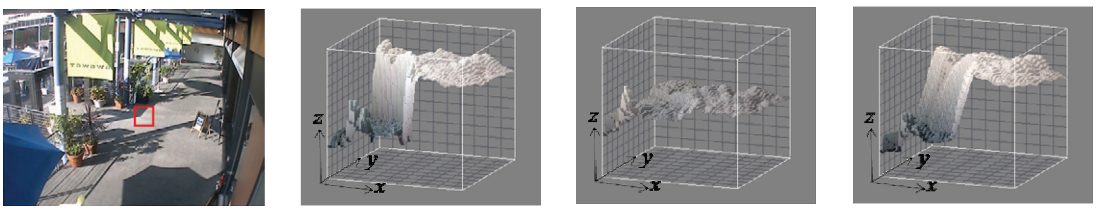


Fig.6. Relation between the differences of the pixel values and the number of the dimensions of the eigenspace.



(i) (ii) (iii) (iv)

Fig.7. The target image and 3D visualization of shadow position of the images. (i) is the target image of the evaluation. The rectangle in the image shows the clipped region. (ii) is the 3D visualization of the clipped region of the target image (i). (iii) and (iv) are the 3D visualization of the generated background image by the simple eigenspace method and proposed method.

Then, we evaluated the accuracy of the background generation in terms of sharpness of the shadow edges. We generated the background image of the target image by the simple eigenspace method and the proposed one. We clipped a region, which included a sharp shadow edge, from the target image. We also clipped the same region from the each generated image. We graphically show the results in Fig.7. The z-axis of each graph indicates the pixel value. Although there is a sharp edge in the target image, no edges are found in the generated image by the simple eigenspace method. On the other hand, in the generated image by the proposed method, we can find the similar sharp edge in the similar position of the target image.

Finally, we evaluated the accuracy of the background generation in terms of the luminosity of the whole image. Fig.8 shows the distribution of the pixel value differences between the target image and generated one. In this graph, the pixel values are sorted in the descending order along the horizontal axis. It is confirmed that the differences of the values are much reduced in the proposed method, which means the proposed method can generate the background image that preserves the illumination condition of the whole image more accurately than the simple method. It is also considered that some large differences still remain even by the proposed method. These differences are considered to be caused by the ghosts mentioned above.

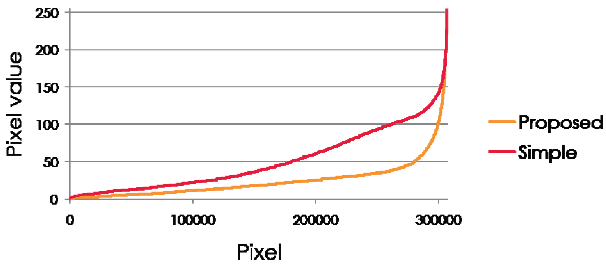


Fig.8. Pixel value differences between the target image and the generated images.

5. Conclusion

In this paper, we proposed a novel background generation technique for a “Henshin” camera, which outputs only privacy-protected information in order to open and share this information in the Web. The information should not contain appearances of people in the captured images but just their positions and the background image of the scene. In addition, it is also required that the generated background has to be as similar to the captured raw image as possible, because a user of the “Henshin” camera would like to recognize the real-time weather or the strength of the sunlight from its background image.

The proposed method is implemented as an extension of the eigenspace method. This method relies on the fact that the lighting conditions of the images which are captured at the same time would be almost the same even if they are captured on different days. To generate the verisimilar background image which well expresses the weather and the lighting condition of the scene, we classified huge number of captured images into the image sets according to time, and then applied the eigenspace method for the images of each set. We experimentally evaluated the results of this approach using data of several surveillance cameras. We achieved better results with the proposed method than the simple eigenspace method without classifying the images into the sets.

Although we can preserve lighting condition of the scene well using our method, we still cannot generate appropriately parking cars which parks for a long time or signage which replace day by day. Our future work contains challenge to make improvements for such background variations. It is also an important topic to determine the suitable term for the proposed method. In this paper, we experimentally determine this term as one month. Dealing with the images observed in very long term is good for the background generation in terms of providing many background variations for generating the background eigenspace. However, it is not good in terms of dealing with the moving shadow edges, because when we collect images observed in very long term, even if they are observed in the same time of different days, positions of the shadows in the images are different caused by seasonal change.

References

Chen, Y.T., & Chen, C.S. (2008) . Fast human detection using a novel boosted cascading structure with meta stages. *IEEE Transactions on Image Processing, 17, 8*, 1452-1464.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 1*, 886-893.

Google (2008). Google Street View. http://maps.google.com

Haritaoglu, I., Harwood, D., & Davis, L.S. (1999). Hydra: Multiple people detection and tracking using silhouettes. *IEEE Workshop on Visual Surveillance, 0*, 6.

Jabri, S., Duric, Z., Wechsler, H., & Rosenfeld, A. (2000). Detection and location of people in video images using adaptive fusion of color and edge information. *International Conference on Pattern Recognition, 4*, 4627.

Minoh, M. (2009). Sensing Web: Concept and Problems (<Special Issue>Sensing Web). [In Japanese], *Journal of Japanese Society for Artificial Intelligence*.

Minoh, M., Kakusho, K., Babaguchi, N., & Ajisaka, T. (2008). Sensing Web Project - How to handle privacy information in sensor data. *Proceedings of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, 863-869.

Oliver, N., Rosario, B., & Pentland, A. (1999). A Bayesian computer vision system for modelling human interactions. *In Proceedings of ICVS99, Gran Canaria, Spain*.